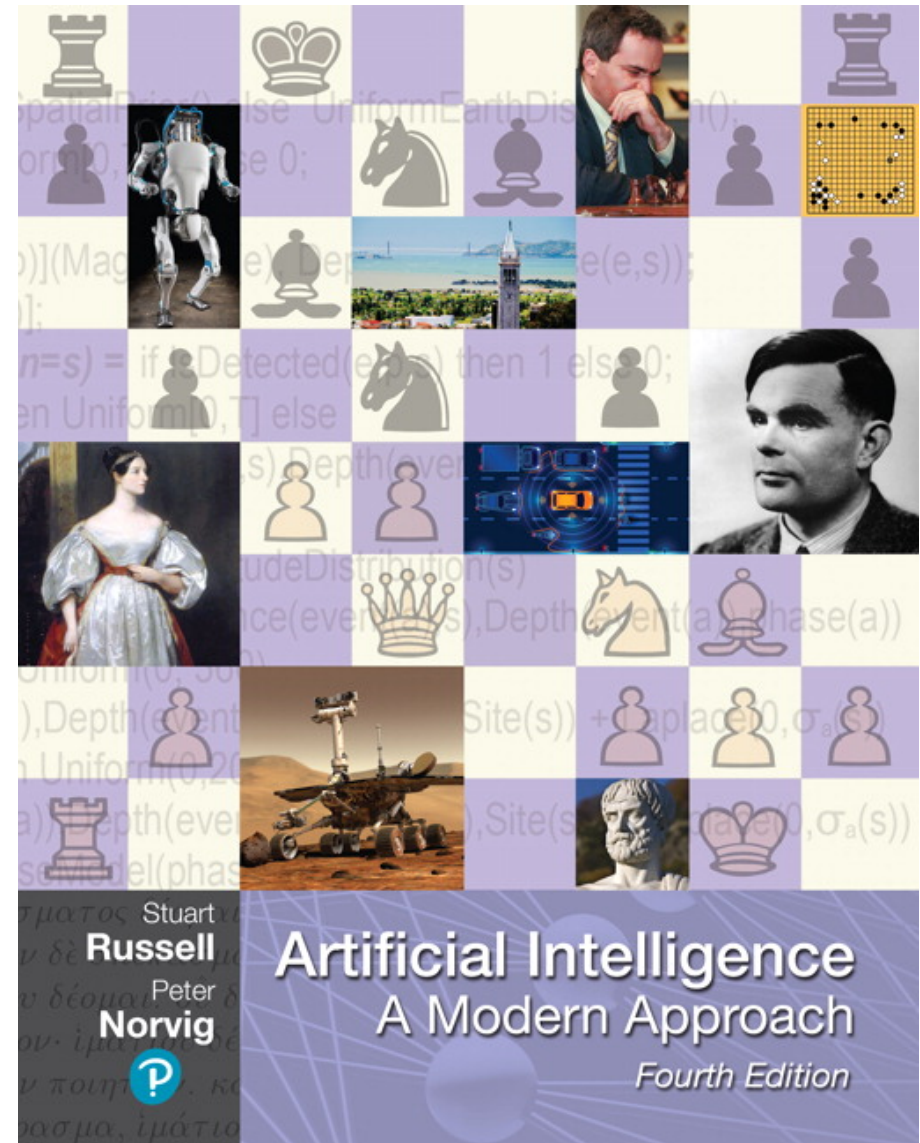


Bayes' Nets

Read AIMA
Chapter 13 "Probabilistic Reasoning"
(Sections 13.1, 13.2 and 13.3)



Slides courtesy of Dan Klein and Pieter Abbeel – University of California, Berkeley

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at <http://ai.berkeley.edu>.]

Probabilistic Models

- Models describe how (a portion of) the world works
- **Models are always simplifications**
 - May not account for every variable
 - May not account for all interactions between variables
 - “All models are wrong; but some are useful.”
– George E. P. Box
- What do we do with probabilistic models?
 - We (or our agents) need to reason about unknown variables, given evidence
 - Example: explanation (diagnostic reasoning)
 - Example: prediction (causal reasoning)



Review: Independence

- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

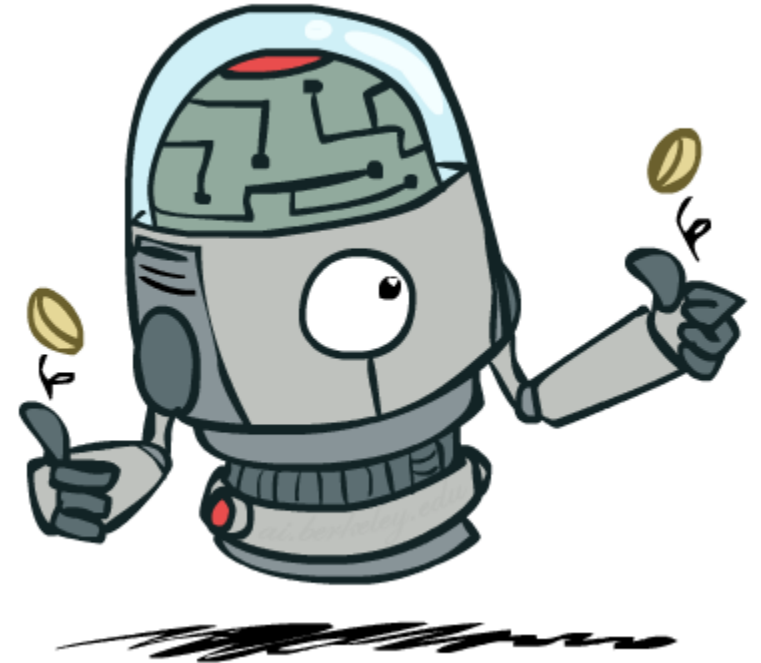
- This says that their joint distribution *factors* into a product two simpler distributions
- Another form:

$$\forall x, y : P(x|y) = P(x)$$

- We write: $X \perp\!\!\!\perp Y$

- Independence is a simplifying *modeling assumption*

- Empirical* joint distributions: at best “close” to independent
- What could we assume for {Weather, Traffic, Cavity, Toothache}?



Review: Independence?

$P_1(T, W)$

| T | W | P |
|------|------|-----|
| hot | sun | 0.4 |
| hot | rain | 0.1 |
| cold | sun | 0.2 |
| cold | rain | 0.3 |

$P(T)$

| T | P |
|------|-----|
| hot | 0.5 |
| cold | 0.5 |

$P_2(T, W)$

| T | W | P |
|------|------|-----|
| hot | sun | 0.3 |
| hot | rain | 0.2 |
| cold | sun | 0.3 |
| cold | rain | 0.2 |

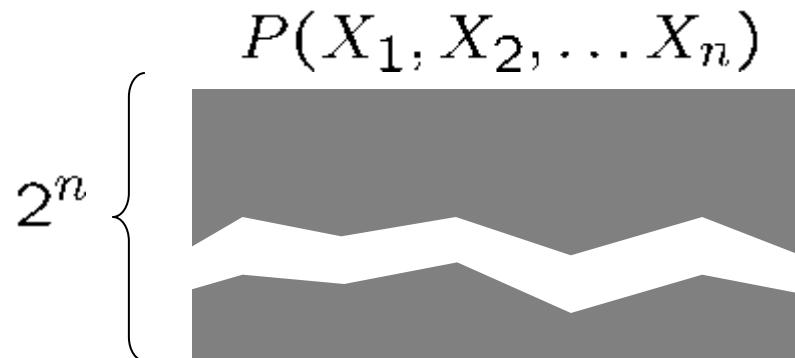
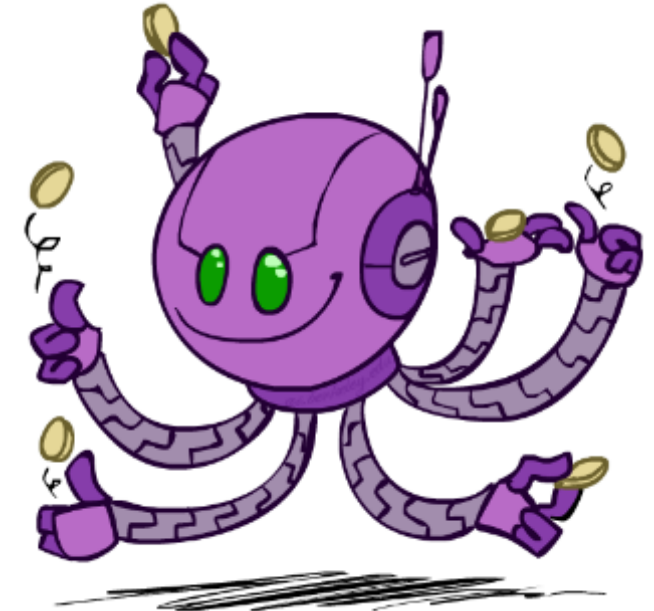
$P(W)$

| W | P |
|------|-----|
| sun | 0.6 |
| rain | 0.4 |

Review: Independence

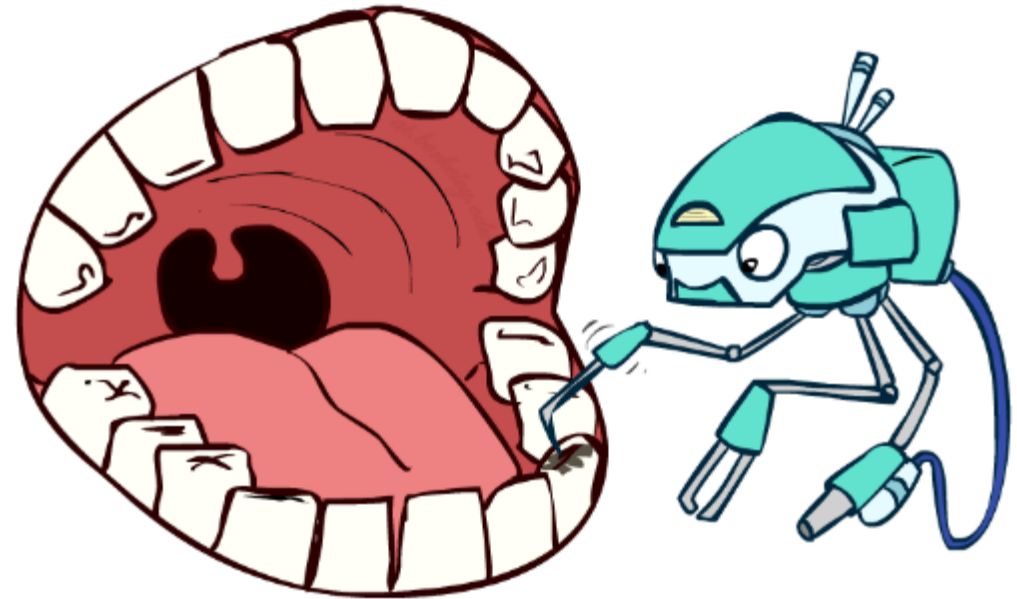
- N fair, independent coin flips:

| $P(X_1)$ | | $P(X_2)$ | | $P(X_n)$ | |
|----------|-----|----------|-----|----------|-----|
| H | 0.5 | H | 0.5 | H | 0.5 |
| T | 0.5 | T | 0.5 | T | 0.5 |



Review: Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Detect})$
- If I have a cavity, the probability that the probe detects it doesn't depend on whether I have a toothache:
 - $P(+\text{detect} \mid +\text{toothache}, +\text{cavity}) = P(+\text{detect} \mid +\text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(+\text{detect} \mid +\text{toothache}, -\text{cavity}) = P(+\text{detect} \mid -\text{cavity})$
- Detect is *conditionally independent* of Toothache given Cavity:
 - $P(\text{Detect} \mid \text{Toothache}, \text{Cavity}) = P(\text{Detect} \mid \text{Cavity})$
- Equivalent statements:
 - $P(\text{Toothache} \mid \text{Detect}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
 - $P(\text{Toothache}, \text{Detect} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Detect} \mid \text{Cavity})$
 - One can be derived from the other using the chain rule



Review: Conditional Independence

- Unconditional (absolute) independence very rare, and it doesn't help us make inferences about other variables.
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z

$$X \perp\!\!\!\perp Y \mid Z$$

if and only if:

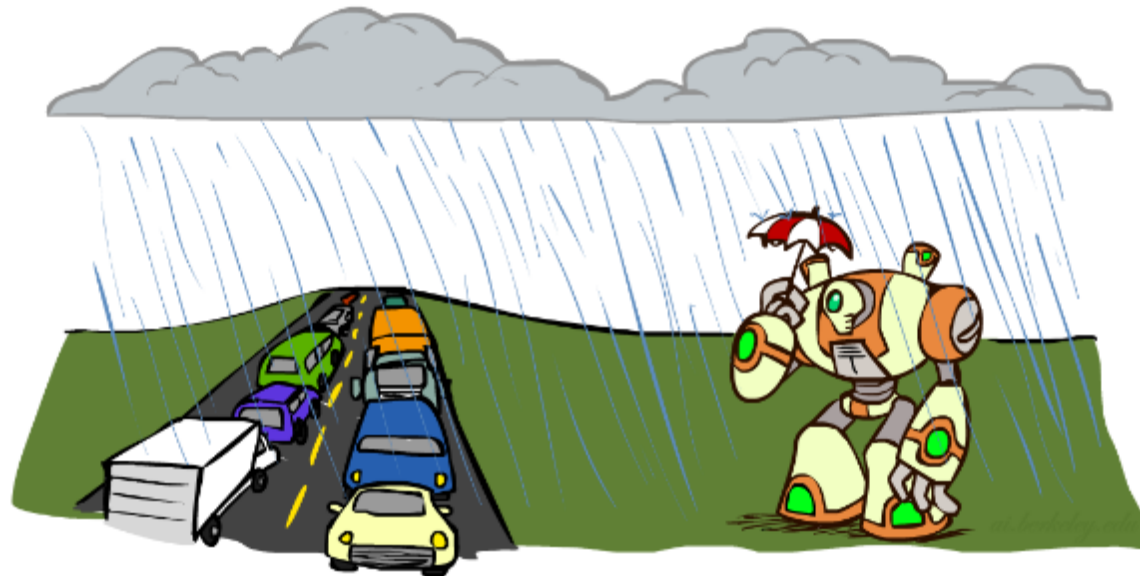
$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

or, equivalently, if and only if

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

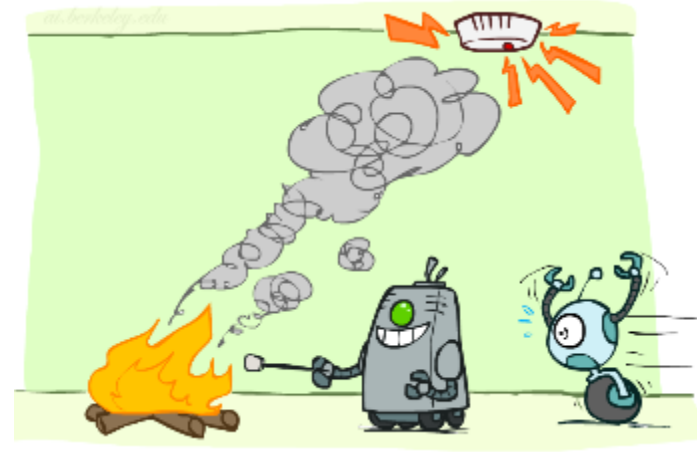
Conditional Independence

- What about this domain:
 - Traffic
 - Umbrella
 - Raining



Conditional Independence

- What about this domain:
 - Fire
 - Smoke
 - Alarm



Conditional Independence and the Chain Rule

- Chain rule: $P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$

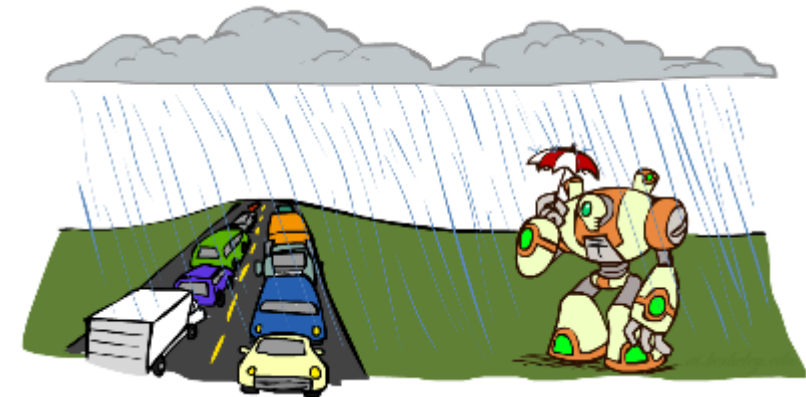
- Trivial decomposition:

$$P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic})$$

- With assumption of conditional independence:

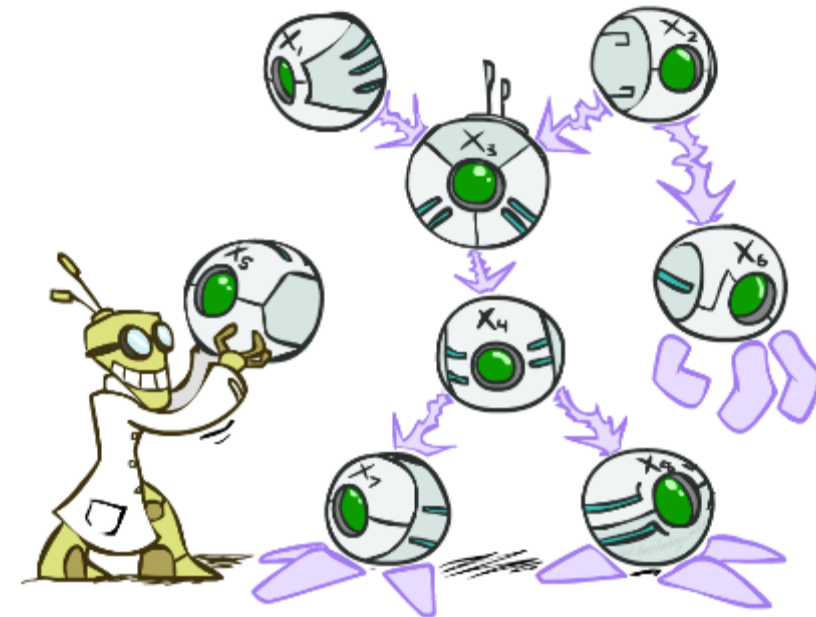
$$P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- Bayes' nets / graphical models help us express conditional independence assumptions

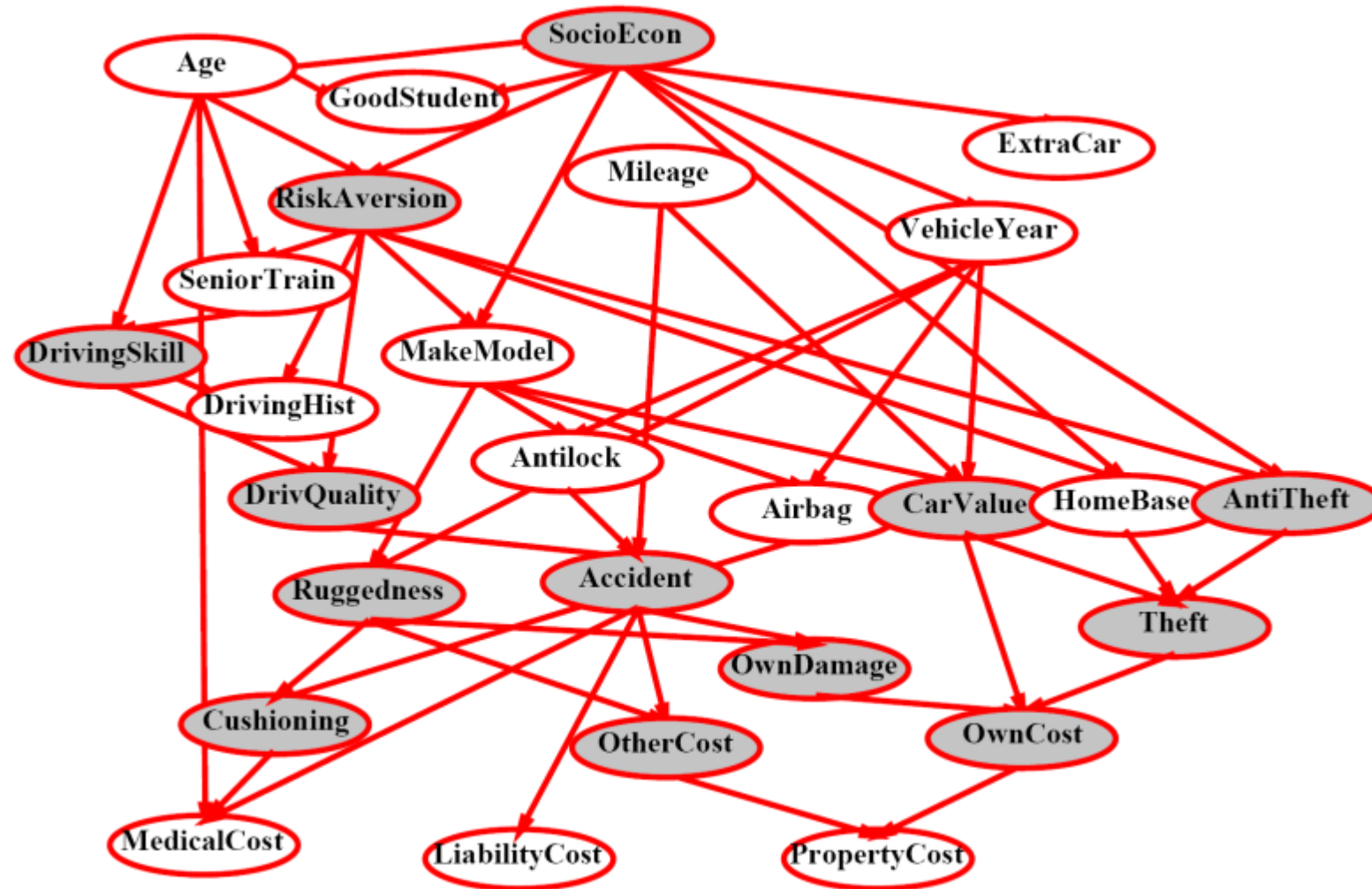


Bayes' Nets: Big Picture

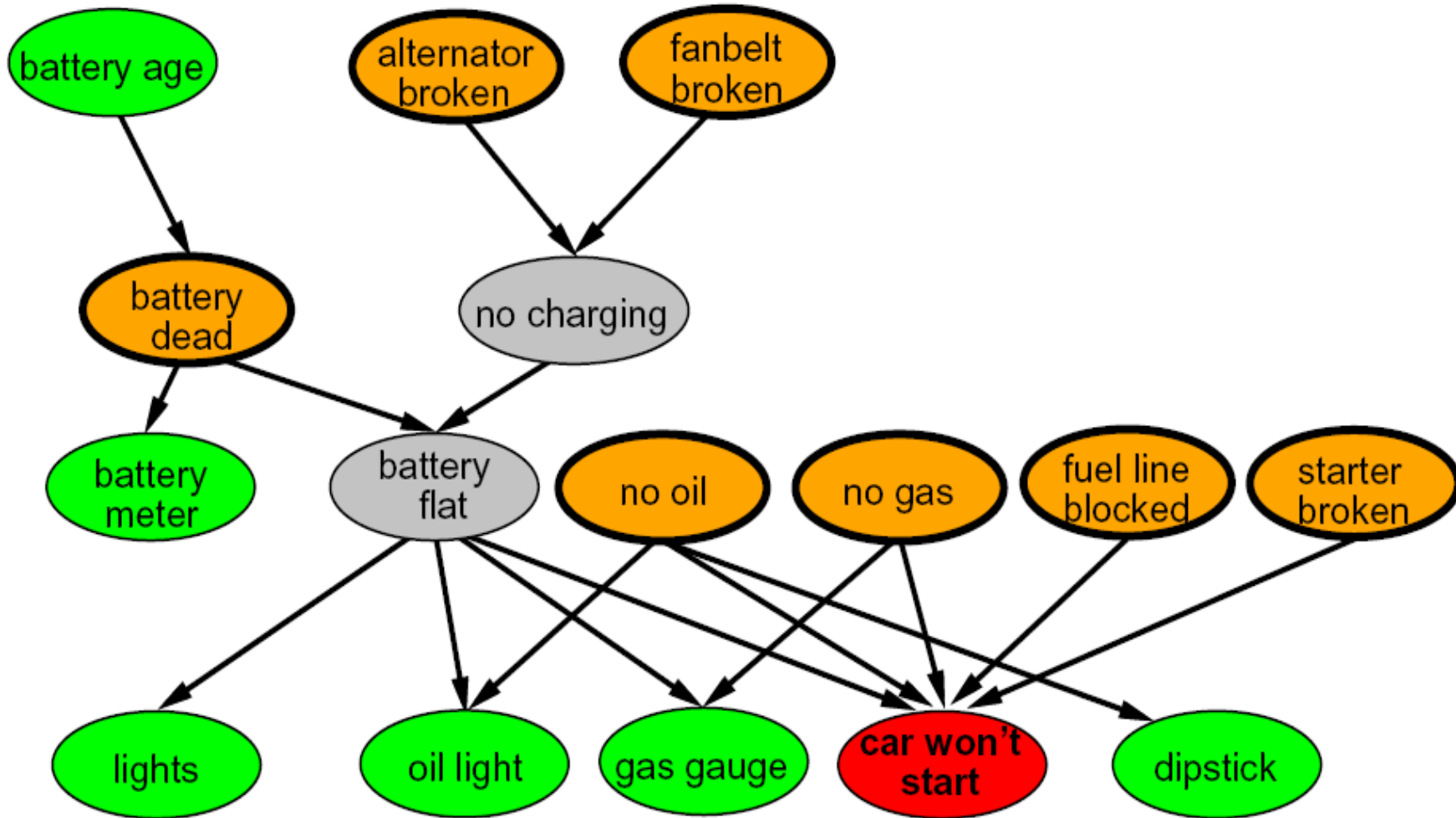
- Two problems with using full joint distribution tables as our probabilistic models:
 - Unless there are only a few variables, the joint is WAY too big to represent explicitly
 - Hard to learn (estimate) anything empirically about more than a few variables at a time
- **Bayes' nets:** a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
 - More properly called **graphical models**
 - We describe how variables locally interact
 - Local interactions chain together to give global, indirect interactions



Example Bayes' Net: Insurance

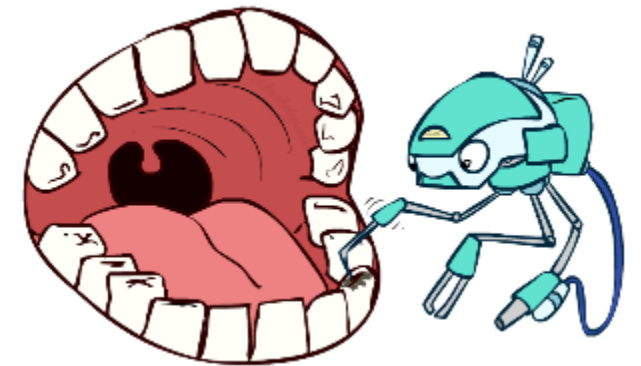
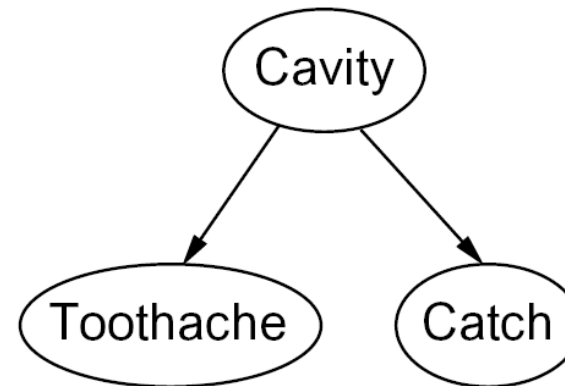
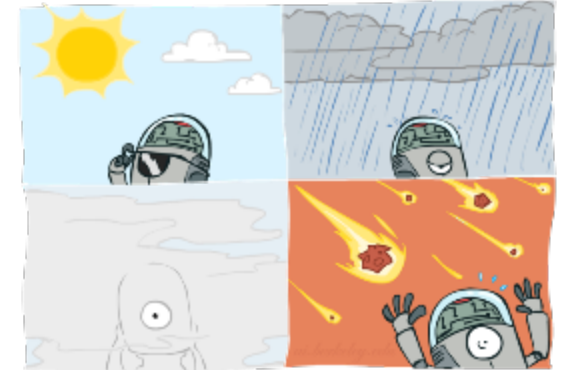


Example Bayes' Net: Car



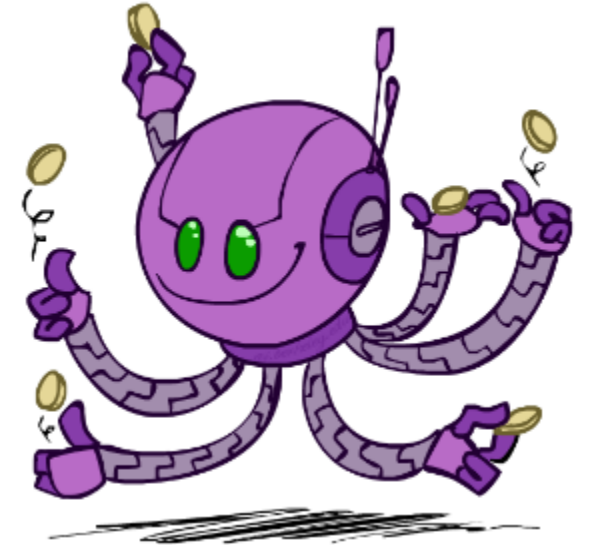
Graphical Model Notation

- **Nodes: variables (with domains)**
 - Can be assigned (observed) or unassigned (unobserved)
- **Arcs: interactions**
 - Indicate “direct influence” between variables
 - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)



Example: Coin Flips

- N independent coin flips



- No interactions between variables: **absolute independence**

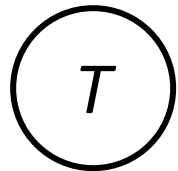
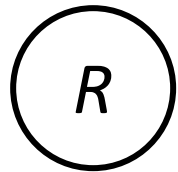
Example: Traffic

- Variables:

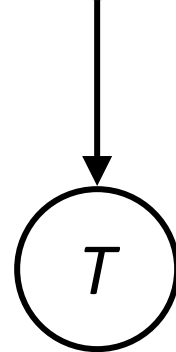
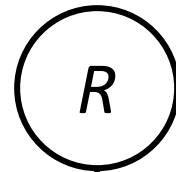
- R: It rains
- T: There is traffic



- Model 1: independence



- Model 2: rain causes traffic



- Why is an agent using model 2 better?

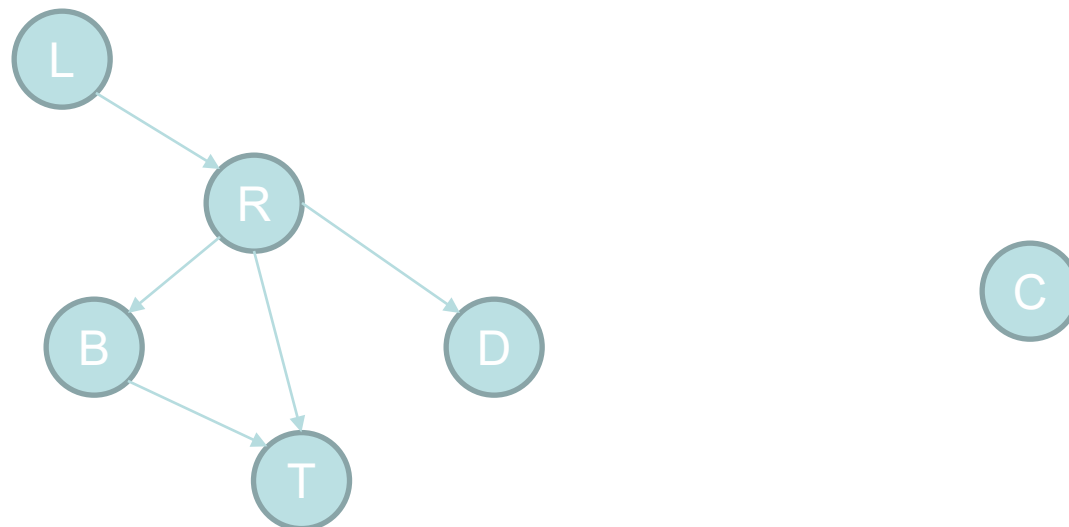
Example: Traffic II

- Let's build a causal graphical model!
- Variables
 - T: Traffic
 - R: It rains
 - L: Low pressure
 - D: Roof drips
 - B: Ballgame
 - C: Cavity



Example: Traffic II

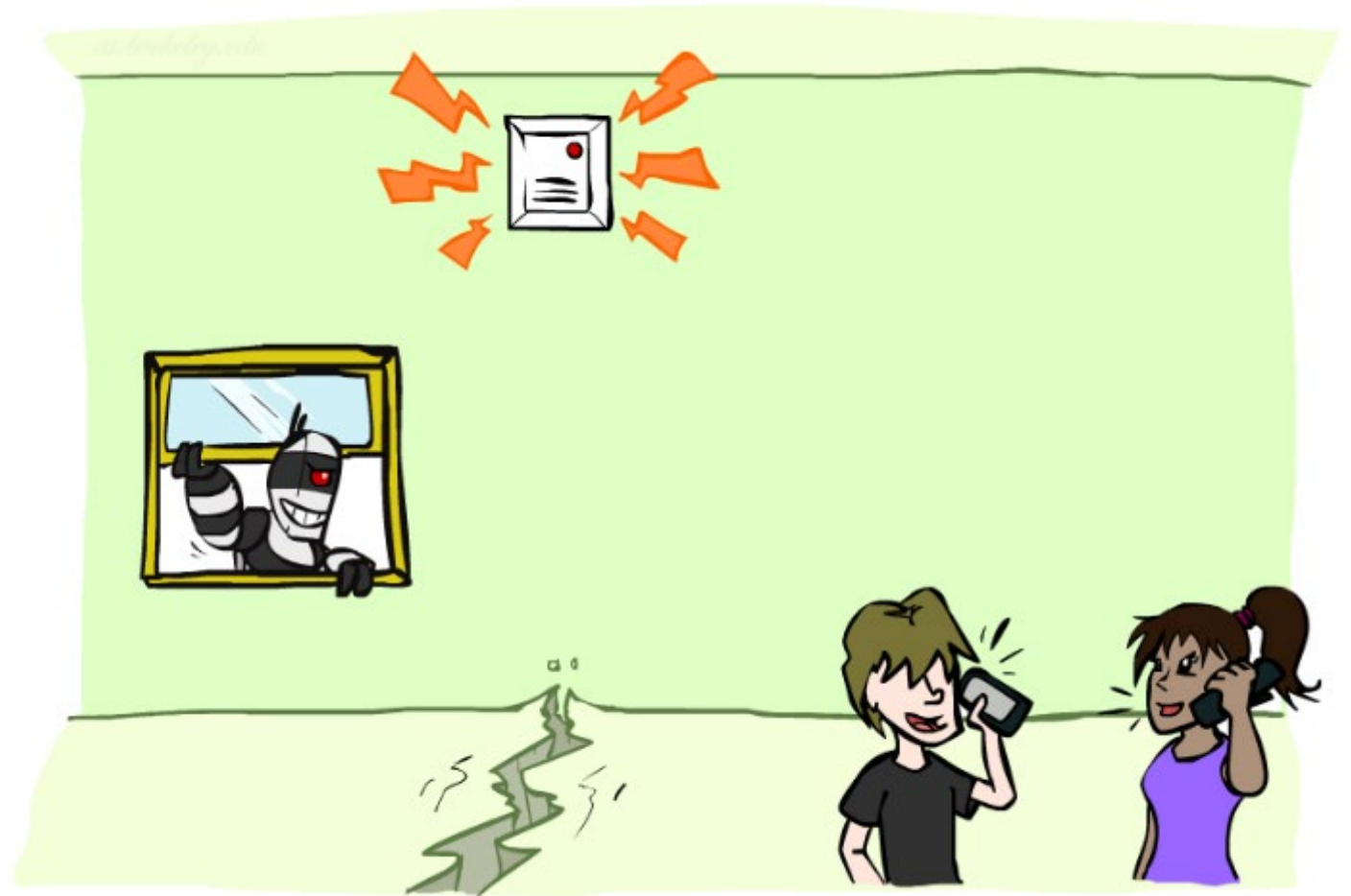
- Let's build a causal graphical model!
- Variables
 - T: Traffic
 - R: It rains
 - L: Low pressure
 - D: Roof drips
 - B: Ballgame
 - C: Cavity



Example: Alarm Network

- Variables

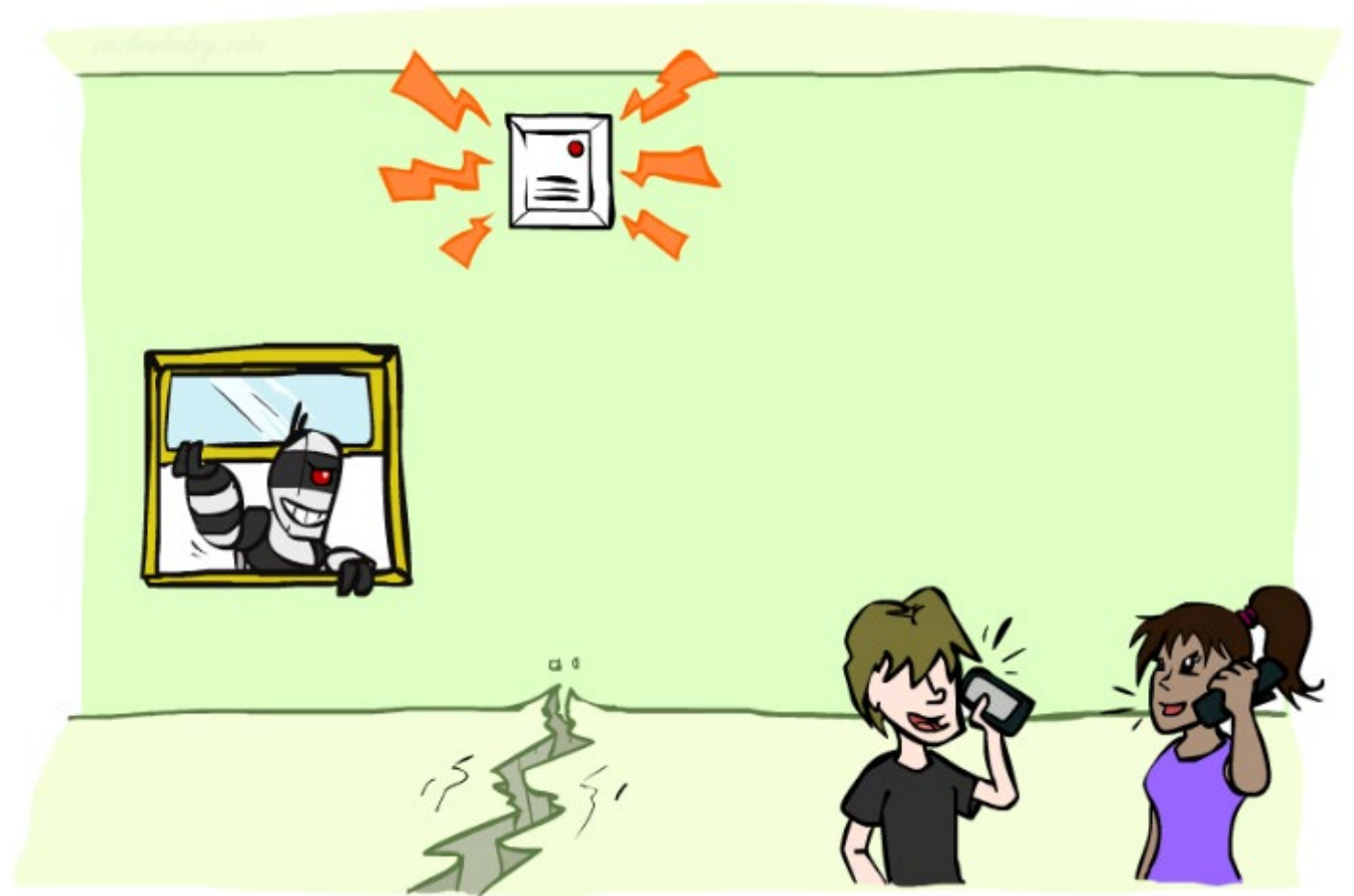
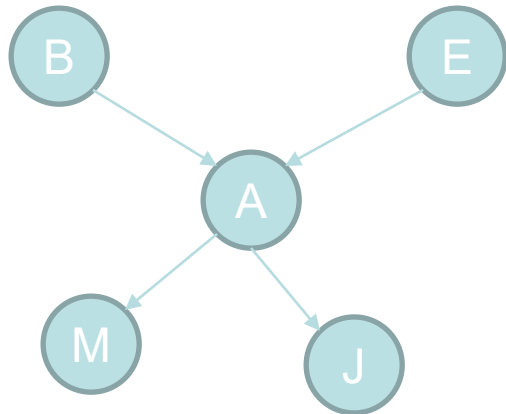
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



Example: Alarm Network

- Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



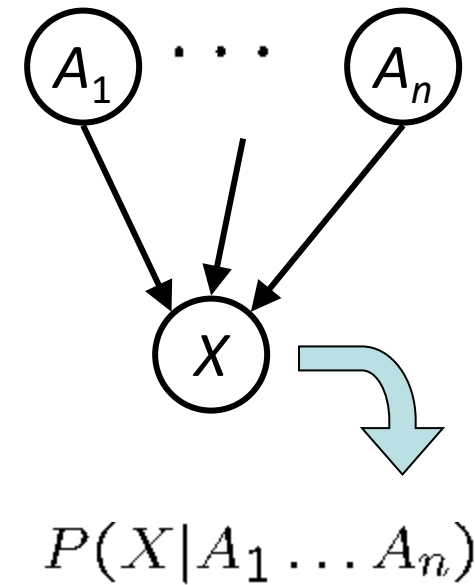
Bayes' Net Semantics



- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
 - A collection of distributions over X , one for each combination of parents' values

$$P(X|a_1 \dots a_n)$$

- CPT: conditional probability table
- Description of a noisy “causal” process



A Bayes net = Topology (graph) + Local Conditional Probabilities

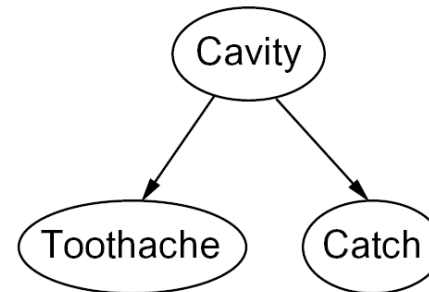
Probabilities in BNs



- Bayes' nets **implicitly** encode joint distributions
 - As a product of local conditional distributions
 - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

- Example:



$$P(+cavity, +detect, -toothache)$$

Probabilities in BNs



- Why are we guaranteed that setting

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

results in a proper joint distribution?

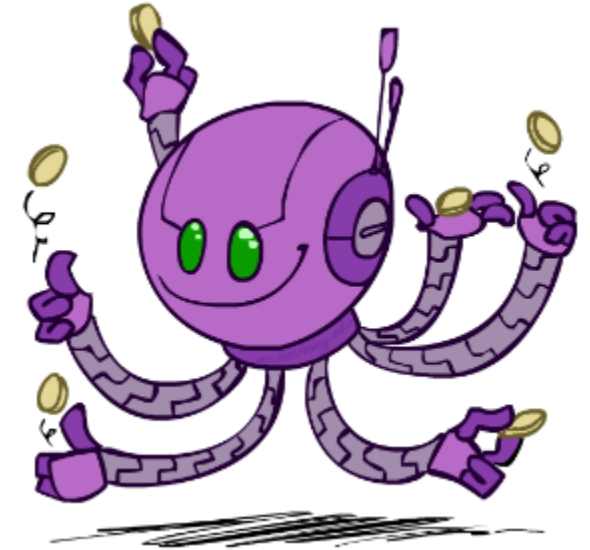
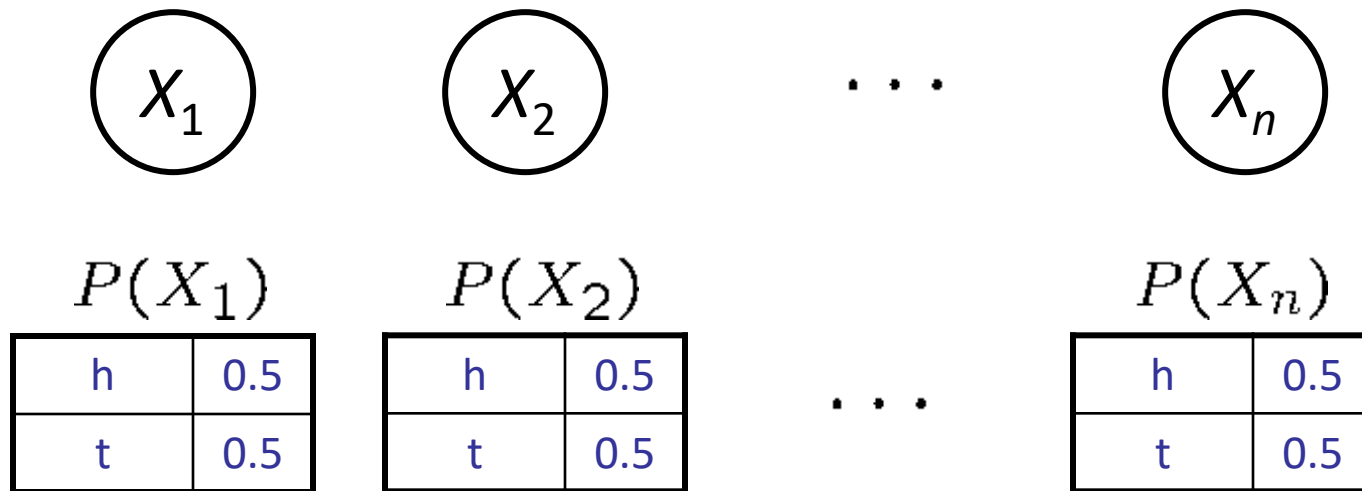
- Chain rule (valid for all distributions): $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1 \dots x_{i-1})$

- Assume conditional independences: $P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$

→ Consequence: $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$

- Not every BN can represent every joint distribution
 - The topology enforces certain conditional independencies

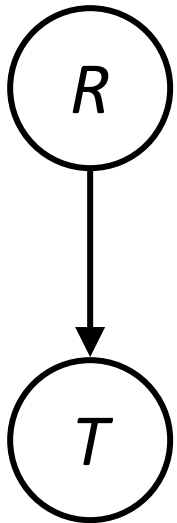
Example: Coin Flips



$$P(h, h, t, h) =$$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic

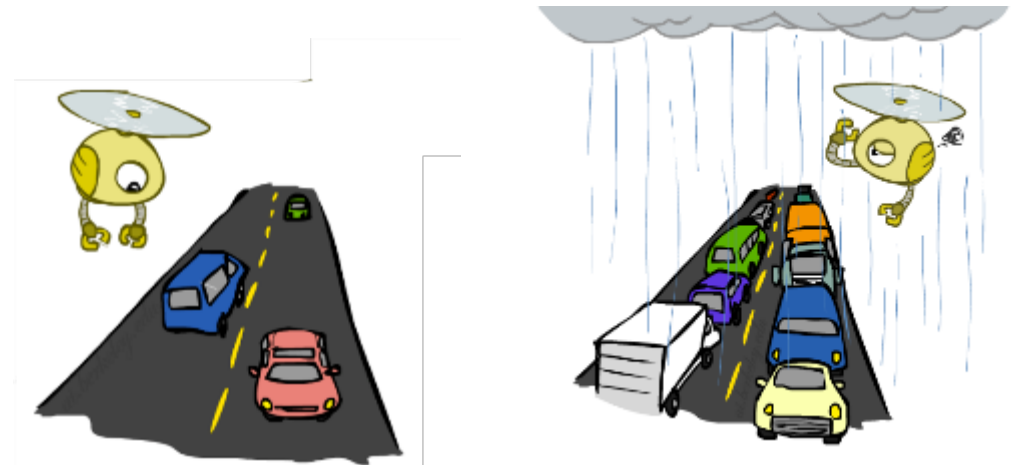

$$P(R)$$

| | |
|----|-----|
| +r | 1/4 |
| -r | 3/4 |

$$P(+r, -t) =$$

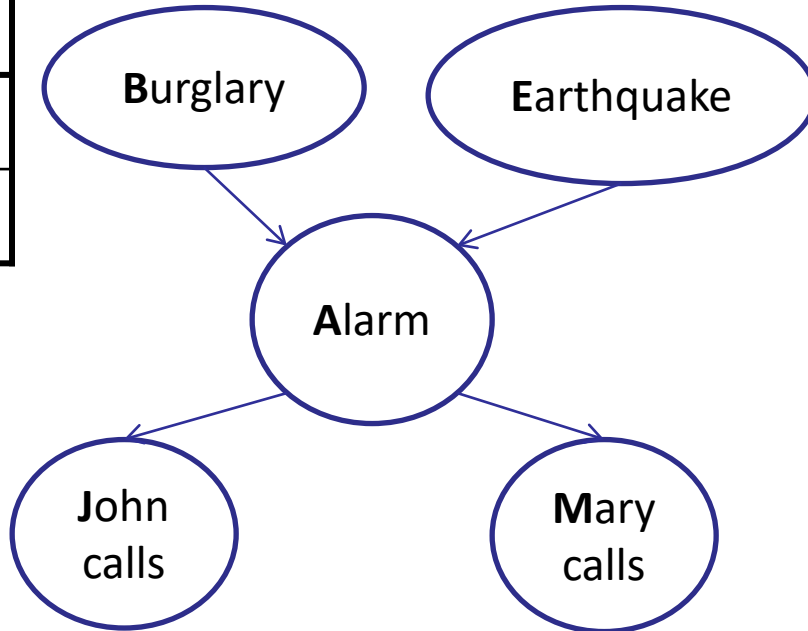
$$P(T|R)$$

| | | |
|----|----|-----|
| +r | +t | 3/4 |
| +r | -t | 1/4 |
| -r | +t | 1/2 |
| -r | -t | 1/2 |

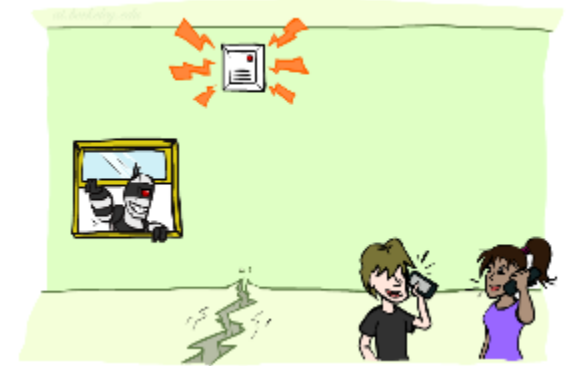


Example: Alarm Network

| B | P(B) |
|----|-------|
| +b | 0.001 |
| -b | 0.999 |



| E | P(E) |
|----|-------|
| +e | 0.002 |
| -e | 0.998 |



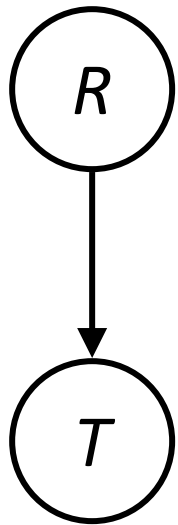
| A | J | P(J A) |
|----|----|--------|
| +a | +j | 0.9 |
| +a | -j | 0.1 |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

| A | M | P(M A) |
|----|----|--------|
| +a | +m | 0.7 |
| +a | -m | 0.3 |
| -a | +m | 0.01 |
| -a | -m | 0.99 |

| B | E | A | P(A B,E) |
|----|----|----|----------|
| +b | +e | +a | 0.95 |
| +b | +e | -a | 0.05 |
| +b | -e | +a | 0.94 |
| +b | -e | -a | 0.06 |
| -b | +e | +a | 0.29 |
| -b | +e | -a | 0.71 |
| -b | -e | +a | 0.001 |
| -b | -e | -a | 0.999 |

Example: Traffic

- Causal direction



$P(R)$

| | |
|----|-----|
| +r | 1/4 |
| -r | 3/4 |

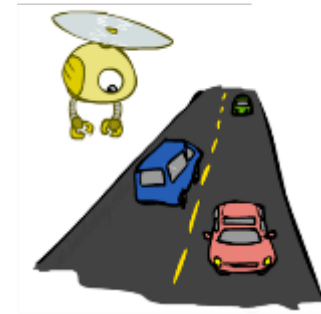
$P(T|R)$

| | | |
|----|----|-----|
| +r | +t | 3/4 |
| | -t | 1/4 |

| | | |
|----|----|-----|
| -r | +t | 1/2 |
| | -t | 1/2 |

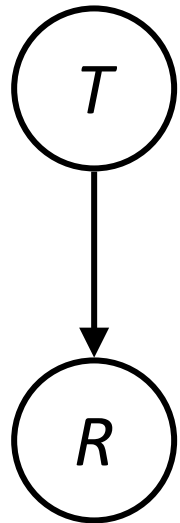
$P(T, R)$

| | | |
|----|----|------|
| +r | +t | 3/16 |
| +r | -t | 1/16 |
| -r | +t | 6/16 |
| -r | -t | 6/16 |



Example: Reverse Traffic

- Reverse causality?



$P(T)$

| | |
|----|------|
| +t | 9/16 |
| -t | 7/16 |

$P(R|T)$

| | | |
|----|----|-----|
| +t | +r | 1/3 |
| | -r | 2/3 |

| | | |
|----|----|-----|
| -t | +r | 1/7 |
| | -r | 6/7 |

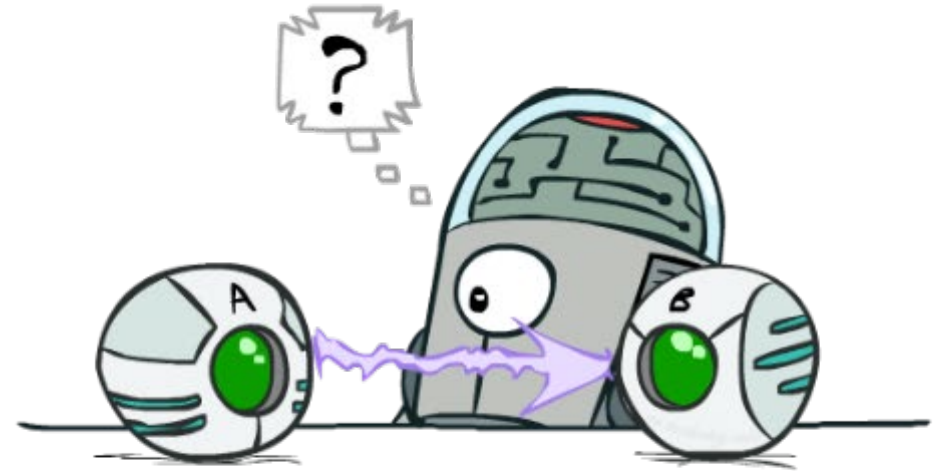


$P(T, R)$

| | | |
|----|----|------|
| +r | +t | 3/16 |
| +r | -t | 1/16 |
| -r | +t | 6/16 |
| -r | -t | 6/16 |

Causality?

- When Bayes' nets reflect the true causal patterns:
 - Often simpler (nodes have fewer parents)
 - Often easier to think about
 - Often easier to elicit from experts
- BNs need not actually be causal
 - Sometimes no causal net exists over the domain (especially if variables are missing)
 - E.g. consider the variables *Traffic* and *Drips*
 - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
 - Topology may happen to encode causal structure
 - **Topology really encodes conditional independence**
$$P(x_i|x_1, \dots, x_{i-1}) = P(x_i|\text{parents}(X_i))$$



Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

$$2^N$$

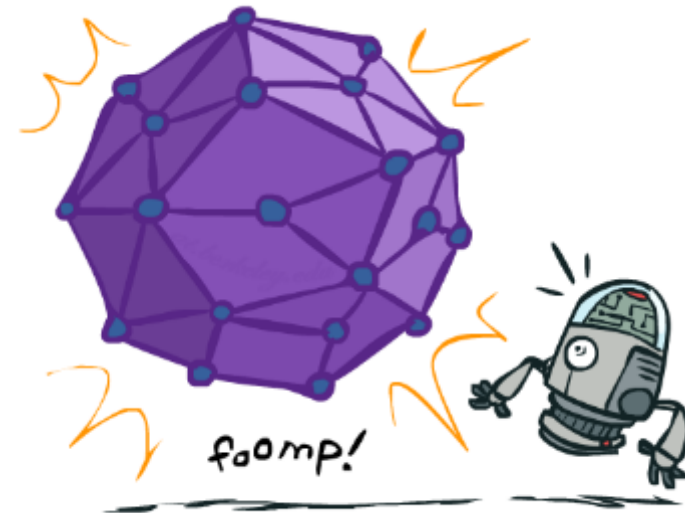
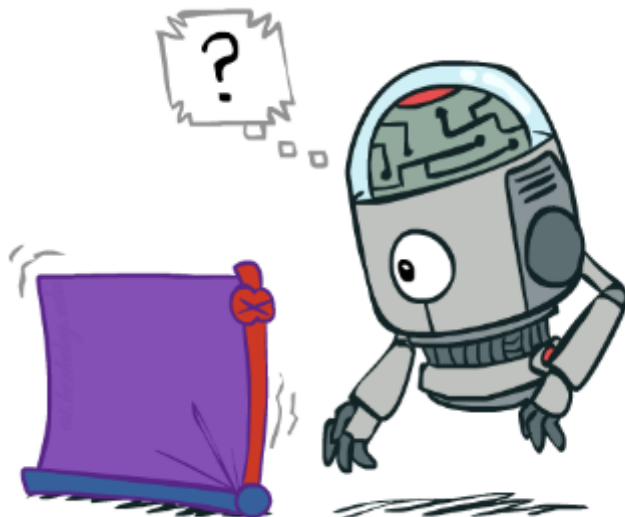
- How big is an N -node net if nodes have up to k parents?

$$O(N * 2^{k+1})$$

- Both give you the power to calculate

$$P(X_1, X_2, \dots, X_n)$$

- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)



Review: Conditional Independence

- X and Y are **independent** if

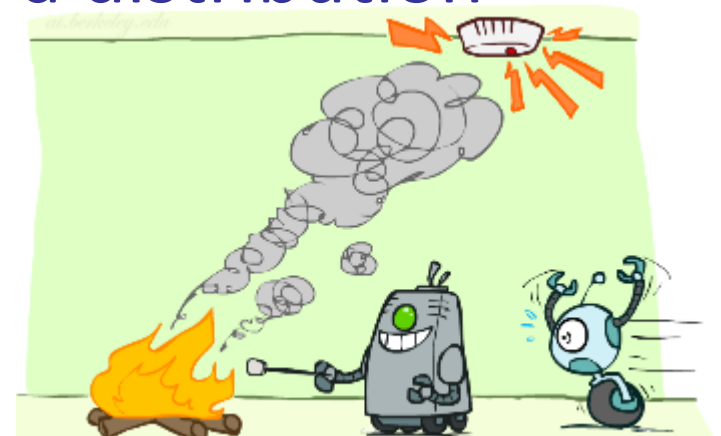
$$\forall x, y \quad P(x, y) = P(x)P(y) \quad \text{---} \rightarrow \quad X \perp\!\!\!\perp Y$$

- X and Y are **conditionally independent** given Z

$$\forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \quad \text{---} \rightarrow \quad X \perp\!\!\!\perp Y|Z$$

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp\!\!\!\perp Fire|Smoke$



Review: Conditional Independence

- Unconditional (absolute) independence very rare, and it doesn't help us make inferences about other variables.
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z

$$X \perp\!\!\!\perp Y \mid Z$$

if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

or, equivalently, if and only if

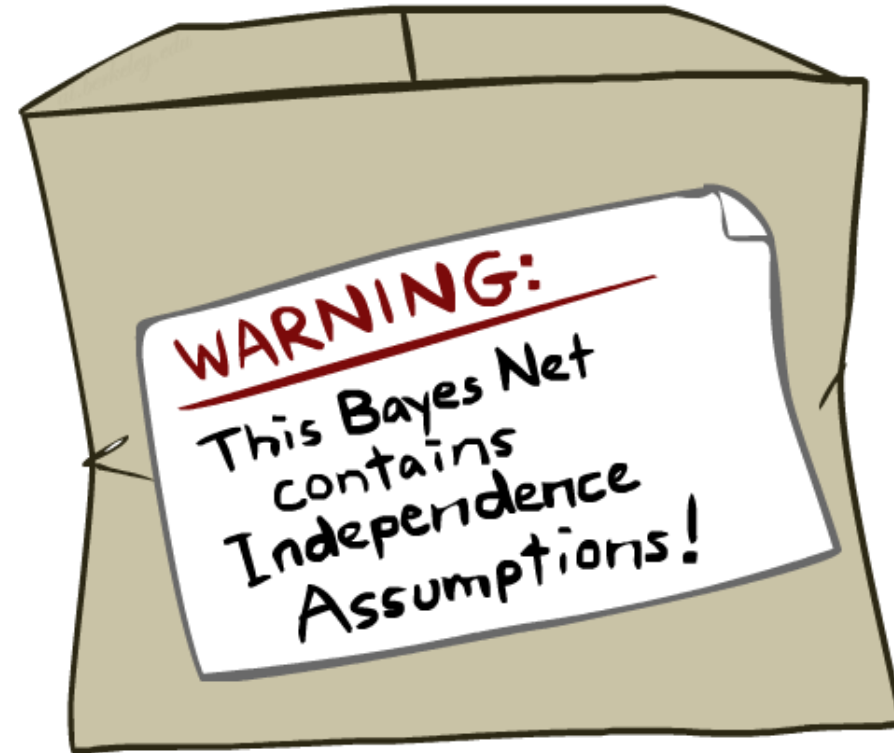
$$\forall x, y, z : P(x|z, y) = P(x|z)$$

Review: Bayes Nets Assumptions

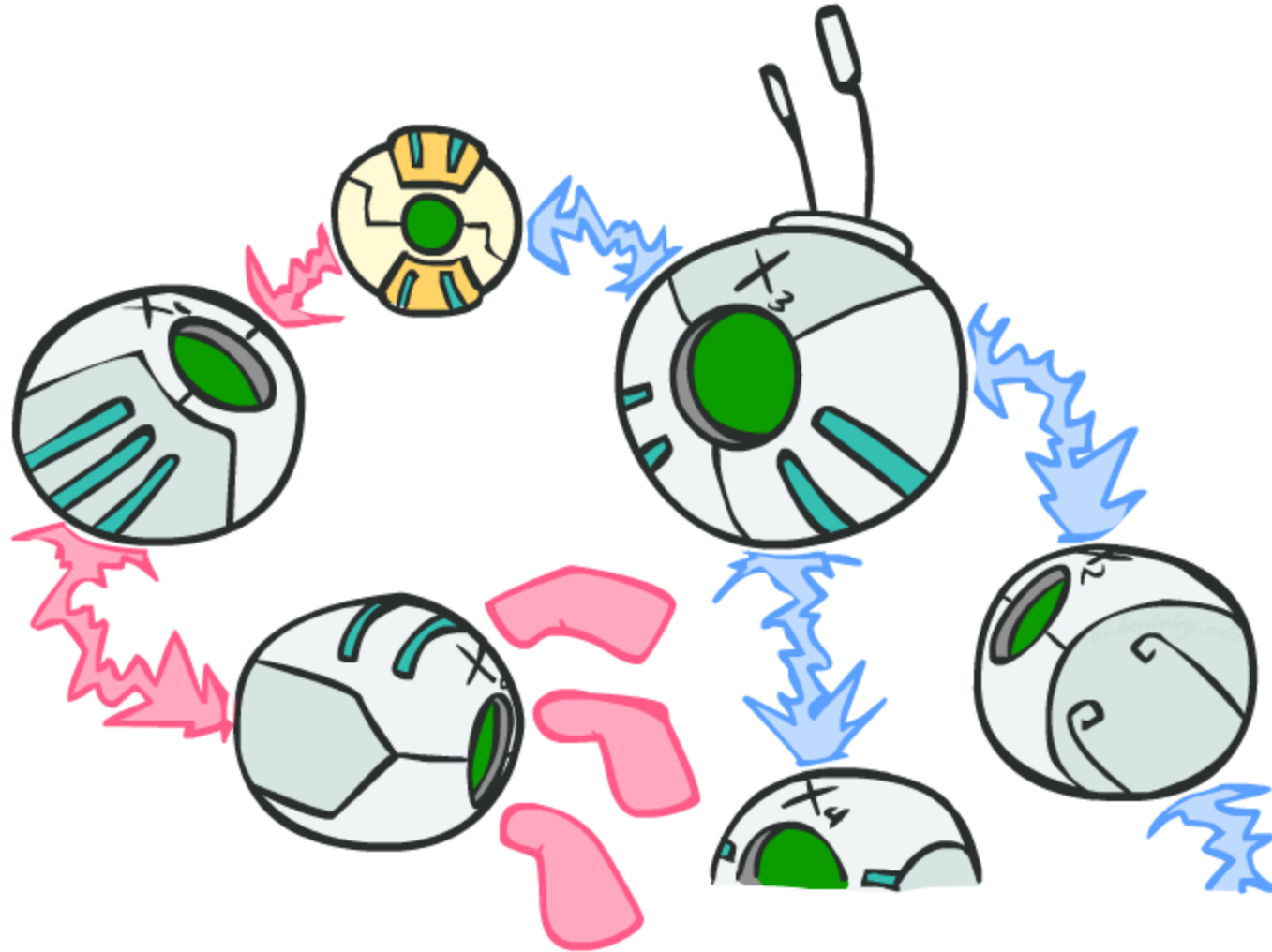
- Assumptions we are required to make to define the Bayes net when given the graph:

$$P(x_i | x_1 \cdots x_{i-1}) = P(x_i | \text{parents}(X_i))$$

- Beyond the “chain rule \rightarrow Bayes net” conditional independence assumptions
 - There are often additional conditional independences
 - They can be read off the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph



D-separation: Outline



Causal Chains

- This configuration is a “causal chain”



X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z ? *No!*

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:

- Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic

- In numbers:

$$P(+y \mid +x) = 1, P(-y \mid -x) = 1, \\ P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

Causal Chains

- This configuration is a “causal chain”

- Guaranteed X independent of Z given Y?



X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

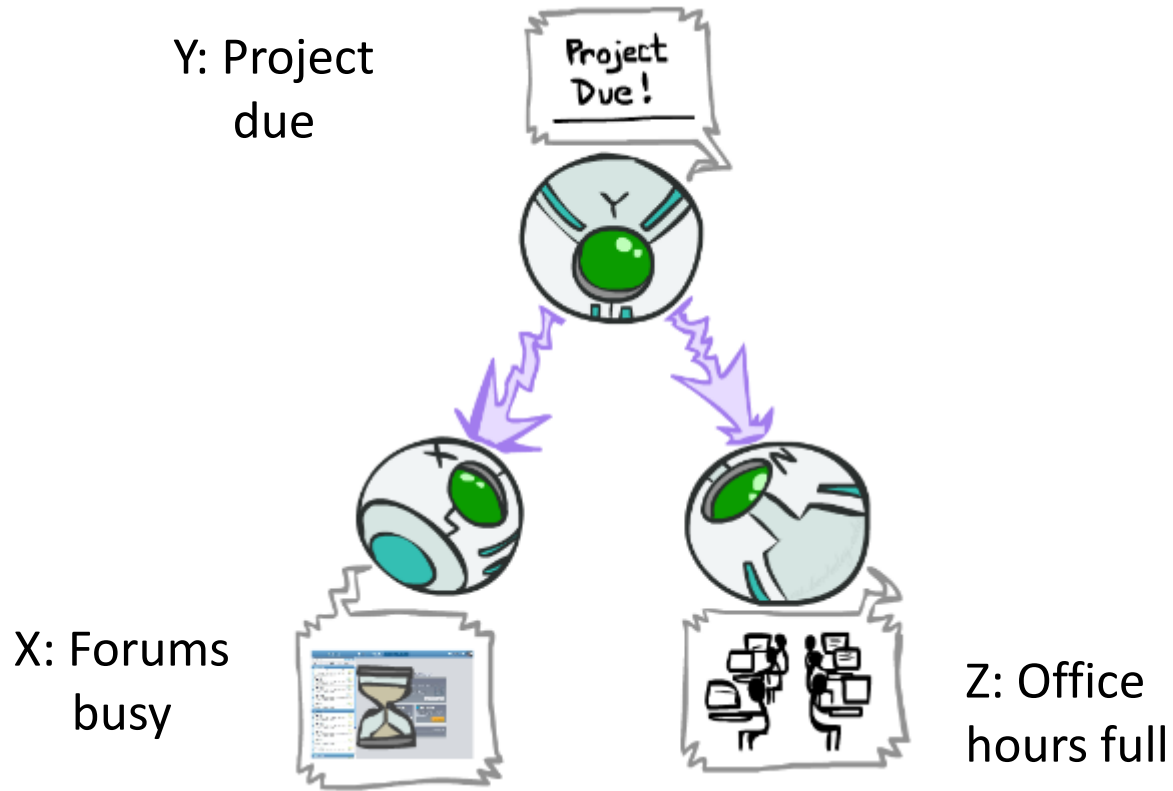
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\ &= P(z|y) \end{aligned}$$

Yes!

- Evidence along the chain “blocks” the influence

Common Cause

- This configuration is a “common cause”



$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X independent of Z ? *No!*

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:

- Project due causes both forums busy and office hours to be full

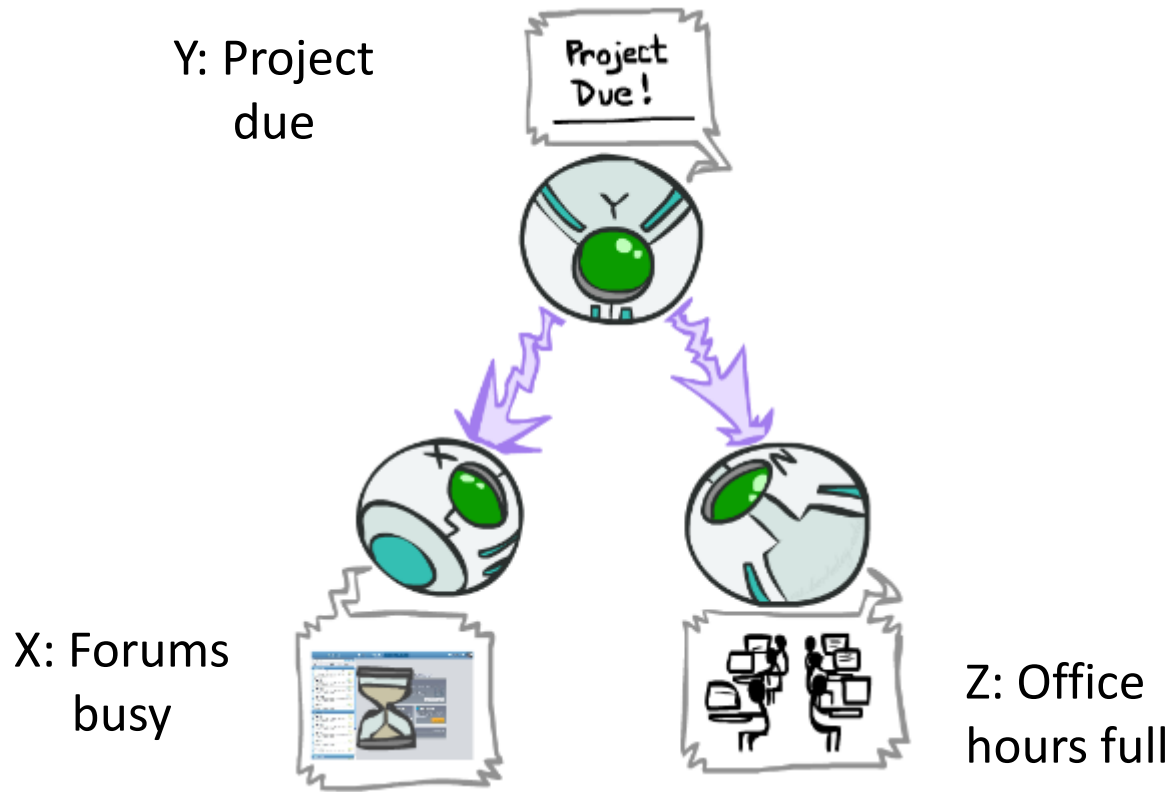
- In numbers:

$$P(+x | +y) = 1, P(-x | -y) = 1, \\ P(+z | +y) = 1, P(-z | -y) = 1$$

Common Cause

- This configuration is a “common cause”

- Guaranteed X and Z independent given Y?



$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

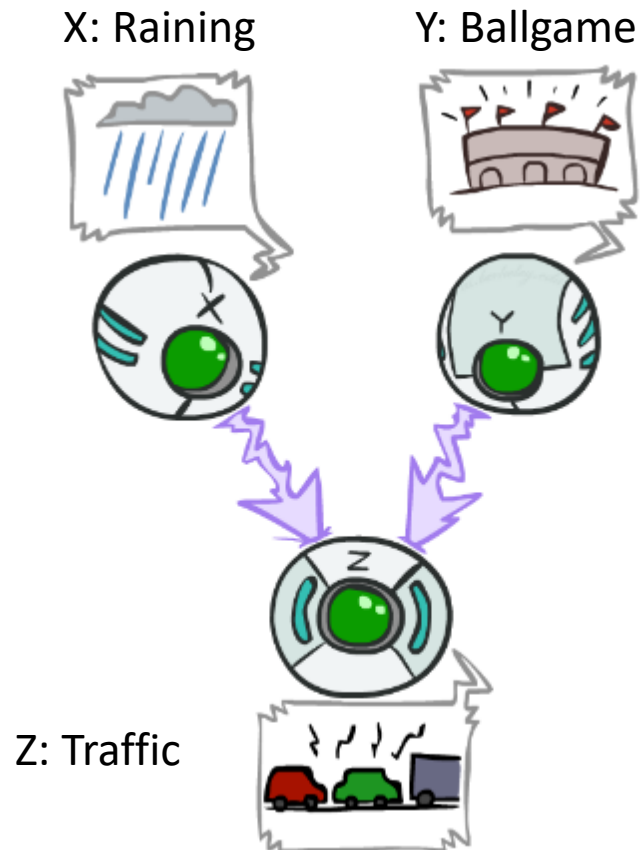
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} \\ &= P(z|y) \end{aligned}$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

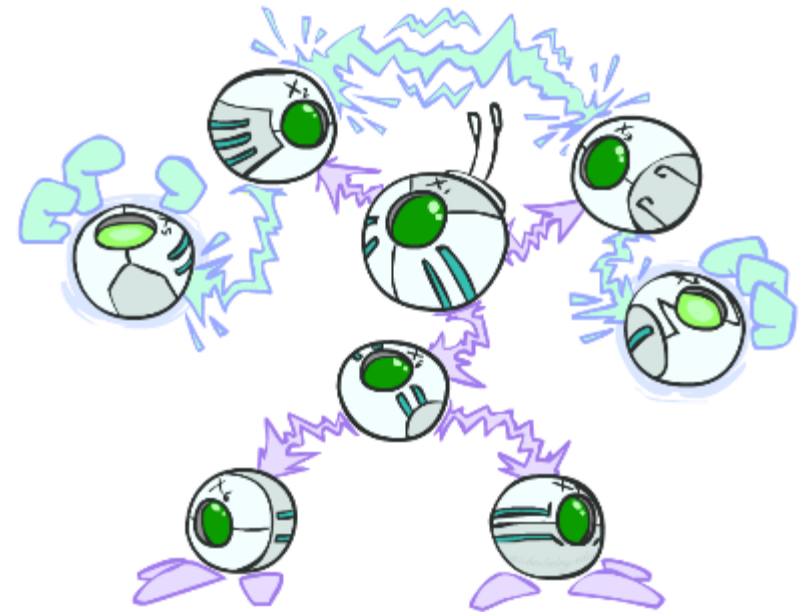
- Last configuration: two causes of one effect (v-structures)



- Are X and Y independent?
 - **Yes**: the ballgame and the rain cause traffic, but they are not correlated
 - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
 - **No**: seeing traffic puts the rain and the ballgame in competition as explanation.
- **This is backwards from the other cases**
 - Observing an effect **activates** influence between possible causes.

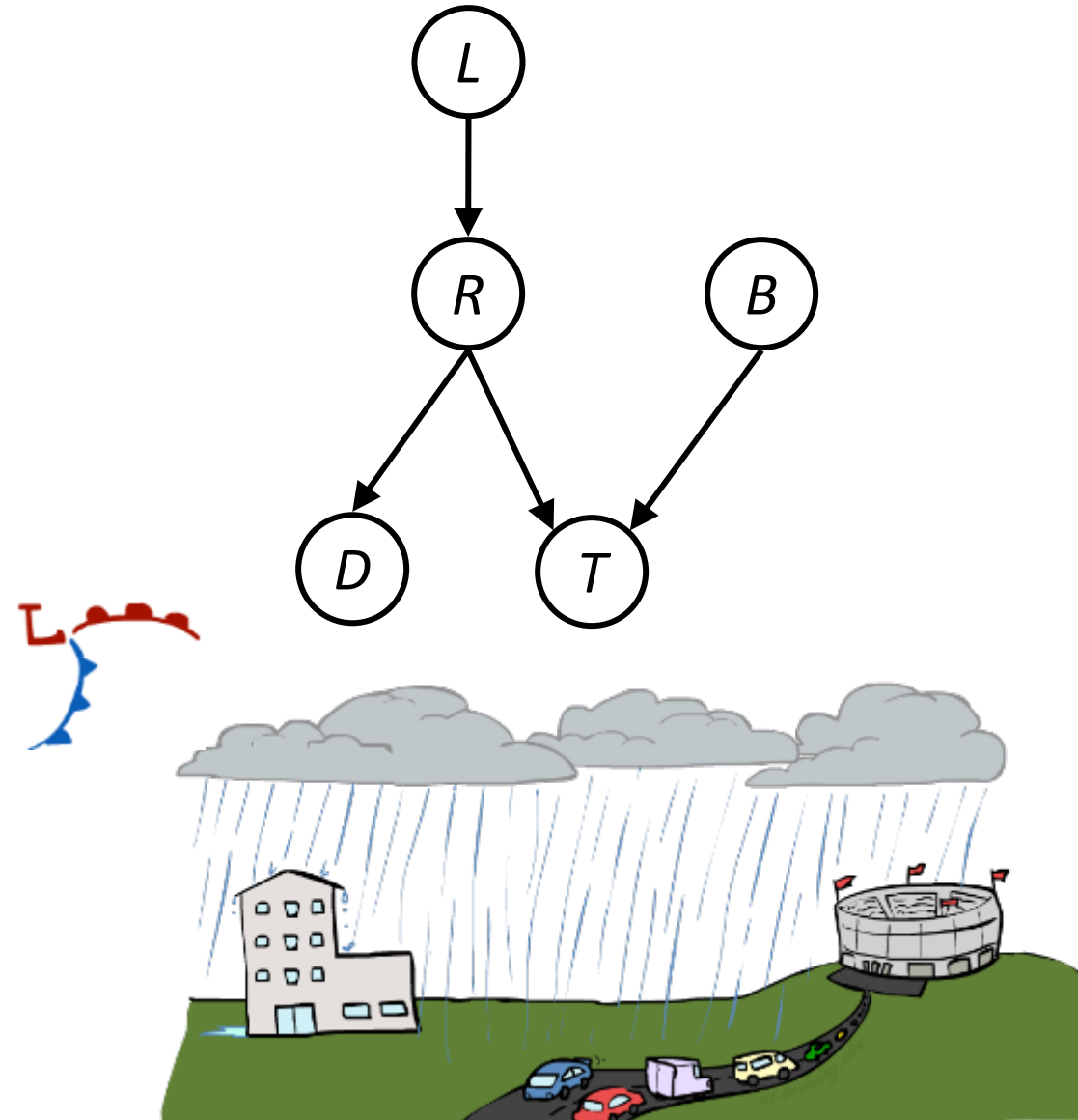
Are two variables in a BN independent?

- General question: in a given BN, are two variables independent (given some evidence)?
- Solution: analyze the graph
- Any complex example can be broken into repetitions of the three canonical cases



Reachability

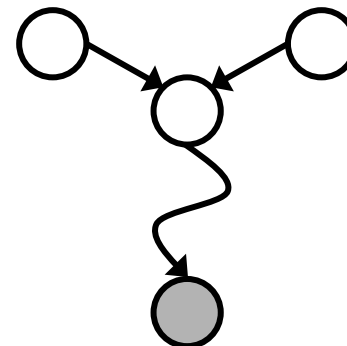
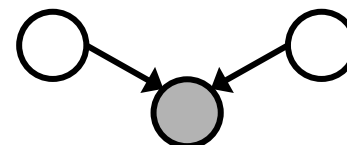
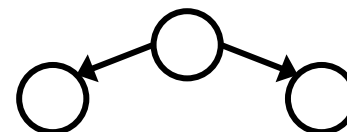
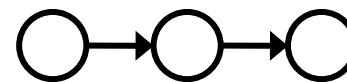
- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
 - Where does it break?
 - Answer: the v-structure at T doesn't count as a link in a path unless "active"



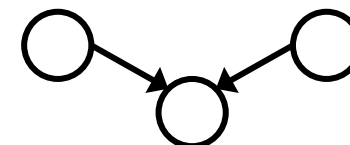
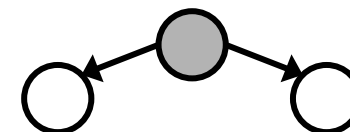
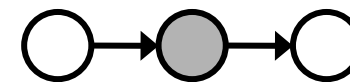
Active / Inactive Paths

- Question: Are X and Y conditionally independent given evidence variables {Z}?
 - Yes, if X and Y “d-separated” by Z
 - Consider all (undirected) paths from X to Y
 - No active paths = independence!
- A path is active if each triple is active:
 - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
 - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
 - Common effect (aka v-structure)
 $A \rightarrow B \leftarrow C$ where B or one of its descendants is observed
- All it takes to block a path is a single inactive segment

Active Triples



Inactive Triples



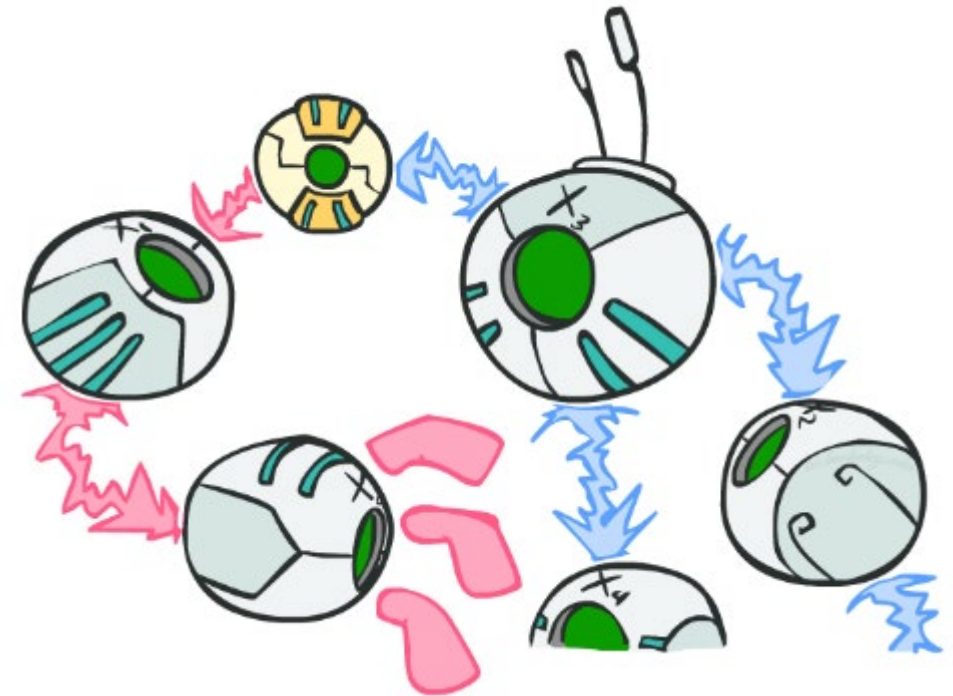
D-Separation

- Query: $X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\} ?$
- Check all (undirected!) paths between X_i and X_j
 - If one or more active, then independence not guaranteed

$$X_i \not\perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

- Otherwise (i.e. if all paths are inactive), then independence is guaranteed

$$X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

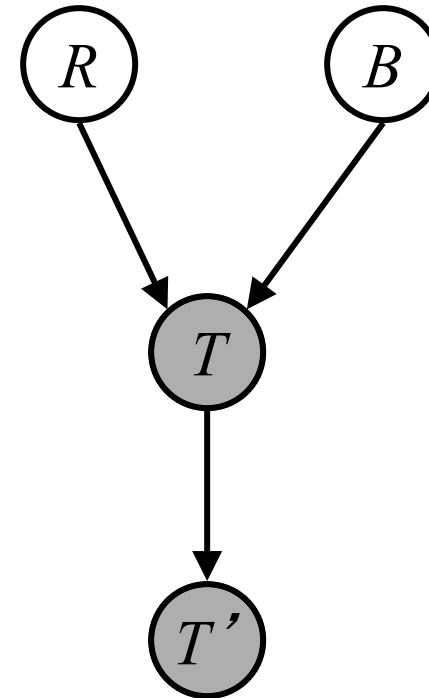


Example

$R \perp\!\!\!\perp B$ *Yes*

$R \perp\!\!\!\perp B | T$ *No*

$R \perp\!\!\!\perp B | T'$ *No*



Example

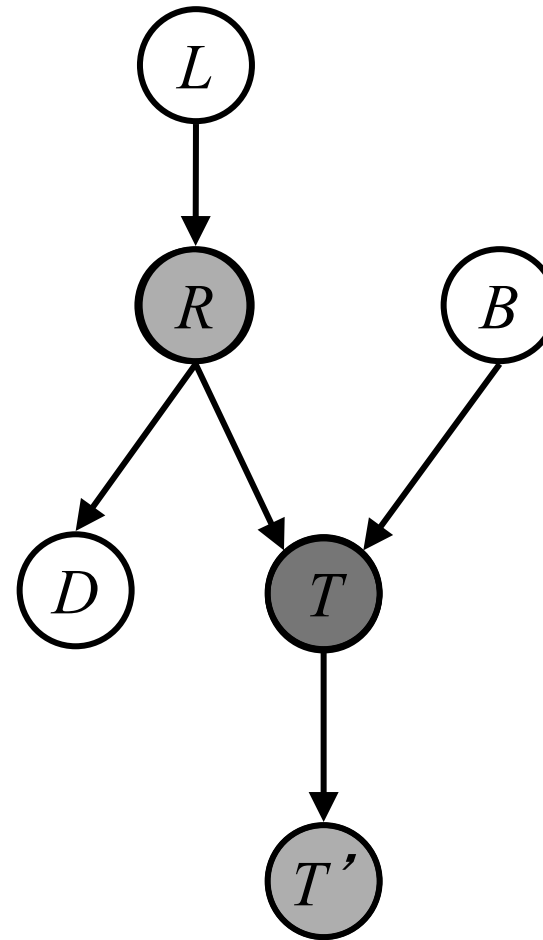
$L \perp\!\!\!\perp T' \mid T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B \mid T$ *No*

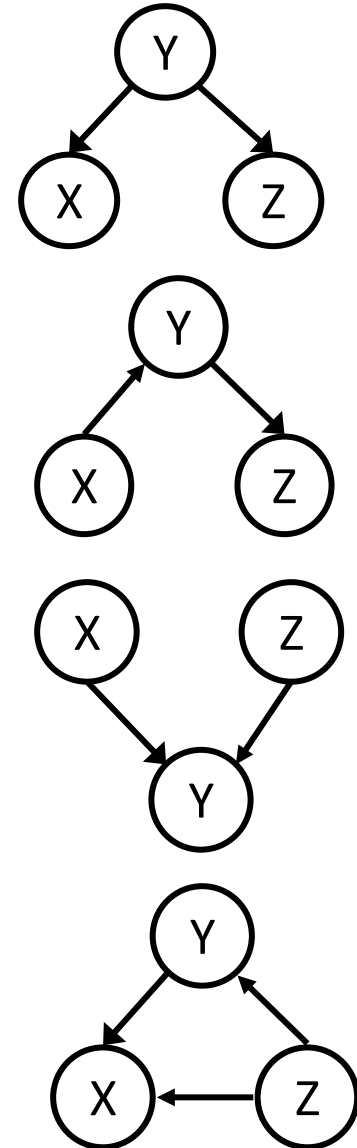
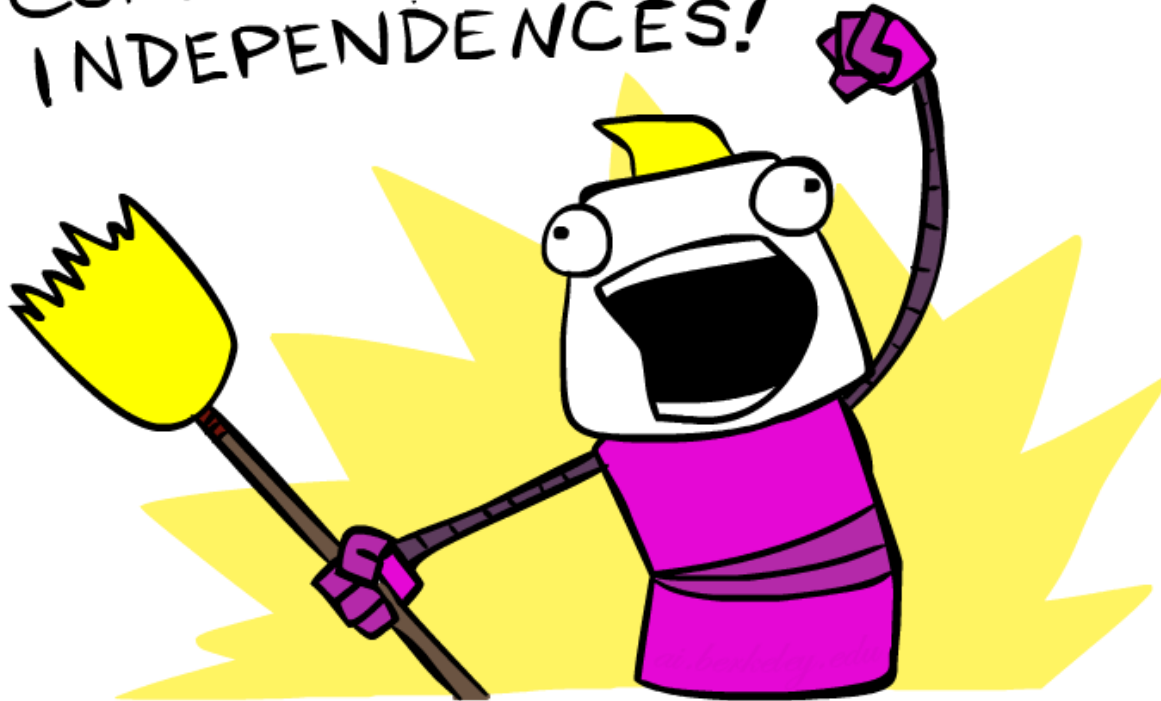
$L \perp\!\!\!\perp B \mid T'$ *No*

$L \perp\!\!\!\perp B \mid T, R$ *Yes*



Computing All Independences

COMPUTE ALL THE
INDEPENDENCES!



Inference

- Inference: calculating some useful quantity from a joint probability distribution

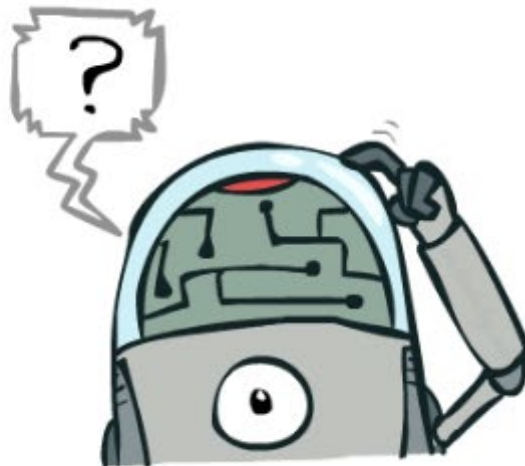
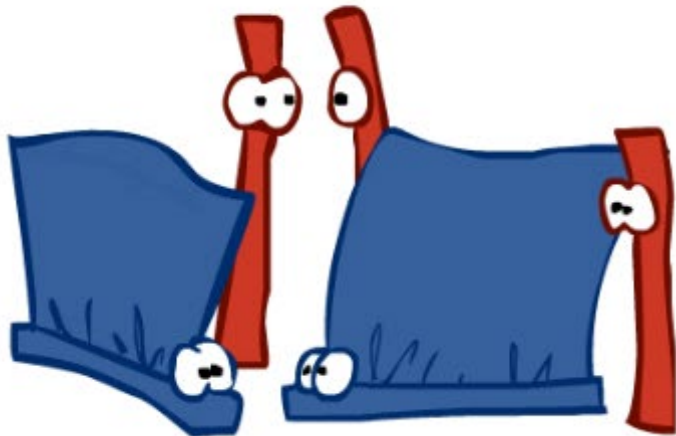
- Examples:

- Posterior probability

$$P(Q|E_1 = e_1, \dots, E_k = e_k)$$

- Most likely explanation:

$$\operatorname{argmax}_q P(Q = q|E_1 = e_1 \dots)$$



Inference by Enumeration

- General case:

- Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
 - Query* variable: Q
 - Hidden variables: $H_1 \dots H_r$
- } X_1, X_2, \dots, X_n
All variables

- We want:

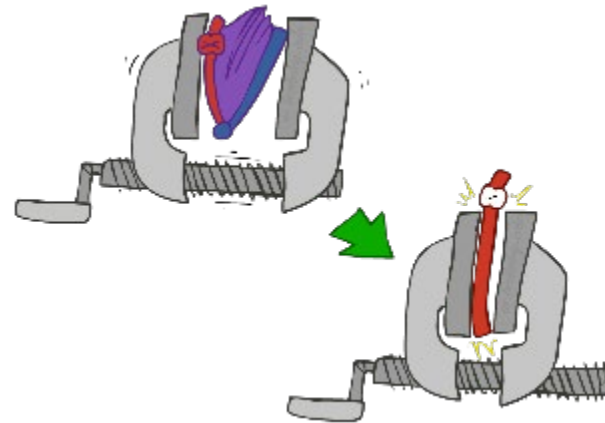
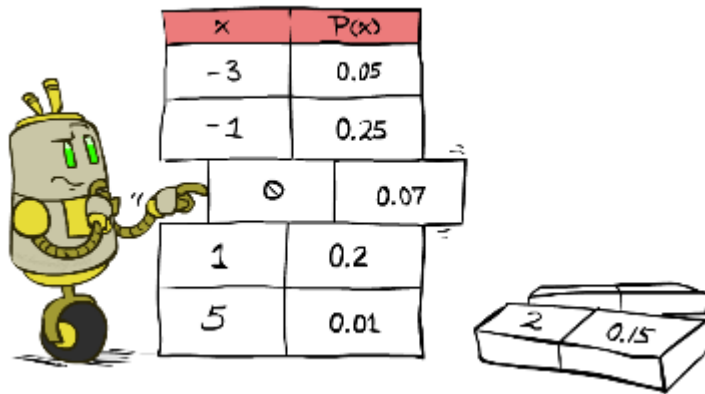
$$P(Q|e_1 \dots e_k)$$

** Works fine with multiple query variables, too*

- Step 1: Select the entries consistent with the evidence

- Step 2: Sum out H to get joint of Query and evidence

- Step 3: Normalize



$$\times \frac{1}{Z}$$

$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots, X_n}$$

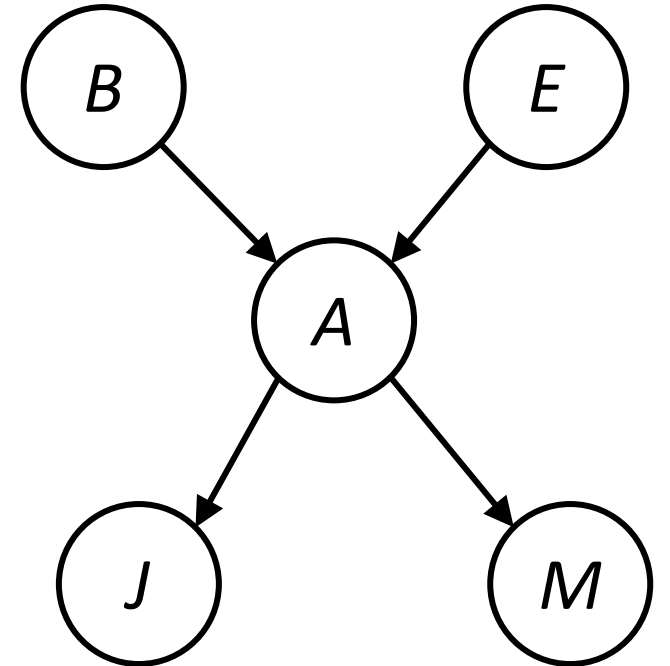
$$Z = \sum_q P(Q, e_1 \dots e_k)$$

$$P(Q|e_1 \dots e_k) = \frac{1}{Z} P(Q, e_1 \dots e_k)$$

Inference by Enumeration in Bayes' Net

- Given unlimited time, inference in BNs is easy
- Reminder of inference by enumeration by example:

$$\begin{aligned}
 P(B \mid +j, +m) &\propto_B P(B, +j, +m) \\
 &= \sum_{e,a} P(B, e, a, +j, +m) \\
 &= \sum_{e,a} P(B)P(e)P(a|B, e)P(+j|a)P(+m|a)
 \end{aligned}$$



$$\begin{aligned}
 &= P(B)P(+e)P(+a|B, +e)P(+j| + a)P(+m| + a) + P(B)P(+e)P(-a|B, +e)P(+j| - a)P(+m| - a) \\
 &\quad P(B)P(-e)P(+a|B, -e)P(+j| + a)P(+m| + a) + P(B)P(-e)P(-a|B, -e)P(+j| - a)P(+m| - a)
 \end{aligned}$$